1. INTRODUCTION

In prediction, linear models of the relationship between predictor and response variables have been dominant in the data sciences and applications. This paper discusses multiplicative models, a class of nonlinear models in which the response is related to the product of the predictors, each raised to a power [1–5]. In contrast to the linear case, in a multiplicative model a fractional increase or decrease in a single predictor produces a proportional change in the response. For example, a company’s sales might be expected to increase proportionally to the size of its sales force. Such multiplicative models have been used to approximate numerous relationships in health [1,6] and labor economics [2], finance [3], environmental chemistry [4], and ecology [7].

The process of fitting a multiplicative model to data is referred to as multiplicative regression. The most common approach is to transform the model into a linear one through a logarithm, followed by linear regression [2,4]. However, when the model is to be used for prediction, it is generally necessary to include and estimate an additional factor accounting for the transformation to avoid a large prediction bias. Several authors [1,2,4,6] have pointed out the importance of this transformation factor and have also suggested that it is often neglected. Alternatively, a multiplicative model may be treated as an instance of a generalized linear model (GLM) [8] with a logarithmic link function and different choices for the relationship between the conditional mean and variance. An empirical comparison of these two approaches was carried out in [2] under a comprehensive range of conditions.

In this paper, we focus on multiplicative modeling for minimum mean squared error (MMSE) prediction. In Section 2, we first study the ideal case in which the model parameters are known and derive the MMSE predictor to establish limits on performance. We show that the coefficient of determination $R^2$, a measure of predictability, is bounded away from its maximum value of 1, with the upper bound depending on the magnitude of the multiplicative noise component. This contrasts with linear prediction, where $R^2$ depends only on the signal-to-noise ratio and is free to approach 1, indicating that multiplicative prediction is a more difficult task.

For the practical case in which the model parameters are estimated from data, we review in Section 3 the methods of ordinary least squares (OLS) on a logarithmic scale and nonlinear least squares (NLS) on the original scale, the latter corresponding to a specific log-link GLM. We then propose a hybrid constrained least squares (CLS) method that minimizes the NLS objective function subject to a confidence region constraint around the OLS solution. In experiments on log-normal and gamma-distributed data, which is a favorable case for log-scale OLS, the proposed CLS method can yield lower mean squared prediction error than OLS when an unmodeled nonlinear component is present, a common occurrence in real-world applications. NLS on the other hand gives inaccurate parameter estimates. For gamma-distributed variables and no model mismatch, both OLS and NLS suffer from occasional large parameter estimation errors, whereas CLS does not and thus performs significantly better. These synthetic experiments suggest that CLS is more flexible than OLS and more stable than NLS in avoiding severe variability. In Section 4.2, we study the performance of the approaches in predicting health care usage and show that CLS continues to outperform log-scale OLS.

2. MODEL AND MINIMUM MEAN SQUARED ERROR PREDICTION

We posit a multiplicative relationship between $p$ non-negative predictor variables $(X_1, \ldots, X_p)$ and a non-negative response variable $Y$:

$$Y = c_0 \prod_{i=1}^{p} X_i^{\beta_i} \epsilon, \quad \epsilon > 0,$$

where $c_0 > 0$ and $\beta_i \in \mathbb{R}$, $i = 1, \ldots, p$ are parameters and $\epsilon$ represents multiplicative error. Taking the logarithm of (1) with $Z = \log Y$, $\beta_0 = \log c_0$, $W_i = \log X_i$, and $\delta = \log \epsilon$ yields the equivalent additive model

$$Z = \beta_0 + \sum_{i=1}^{p} \beta_i W_i + \delta.$$

We make the standard assumptions in regression of mean independence: $E[\epsilon|X] = E[\epsilon] = \mu_\epsilon$ and $E[\delta|W] = E[\delta] = \mu_\delta$, and homoscedasticity: $\text{var}(\epsilon|X) = \text{var}(\epsilon) = \sigma_\epsilon^2$ and $\text{var}(\delta|W) = \text{var}(\delta) = \sigma_\delta^2$. These are all implied if $\epsilon$ is independent of $X$. We then assume without loss of generality that $\mu_\delta = 0$ and $E[W_i] = 0$ for all $i$ since any nonzero means can be absorbed into $\beta_0$. 

**MULTIPLICATIVE REGRESSION VIA CONSTRAINED LEAST SQUARES**

Dennis Wei, Karthikeyan Natesan Ramamurthy, Dmitriy A. Katz-Rogozhnikov, Aleksandra Mojsilović

IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

**ABSTRACT**

This paper considers multiplicative models for predicting a response variable as a product of predictor variables. In the ideal case of known model parameters, the minimum mean squared error predictor is derived and its performance is shown to be fundamentally limited by the magnitude of the multiplicative error component. For estimating model parameters from data, the methods of logarithmically-transformed ordinary least squares (OLS) and nonlinear least squares (NLS) are discussed. We then propose a constrained least squares (CLS) regression method that combines the NLS objective function with a constraint based on the OLS solution. In experiments on log-normal and gamma-distributed data, CLS yields significant improvements in mean squared prediction error by avoiding large errors in parameter estimates and better accommodating model mismatch. We also compare the performances of the regression methods using real-world health care usage data.

**Index Terms—** Predictive models, multiplicative models, regression analysis, least squares methods, health care usage
Our interest is in predictors $\hat{Y}(X)$ of $Y$ with low mean squared error,
\[
\text{MSE} = E \left[ (\hat{Y}(X) - Y)^2 \right].
\]
Common normalized measures of prediction performance are the variance reduction factor $\text{MSE}/\text{var}(Y)$ and its complement, the (population) coefficient of determination $R^2 = 1 - \text{MSE}/\text{var}(Y)$.

As is well known [9], the ideal predictor that minimizes the MSE (3) is the conditional mean $E[Y|X]$. Under the multiplicative model (1), this is given by
\[
E[Y|X] = c_0 \mu_0 \prod_{i=1}^{p} X_{i}^{\beta_i}. \tag{4}
\]
The MMSE predictor is thus scaled by the mean $\mu_0$, in addition to $c_0$, as has been noted e.g. in [1, 2, 4]. In the conventional OLS method discussed in Section 3.1, $c_0$ and $\mu_0$ are estimated separately, whereas in the NLS method in Section 3.2, a single combined scale factor is estimated. An alternative possibility is to assume that $\mu_0 = 1$, in effect absorbing $\mu_0$ into $c_0$, but in that case it cannot also be assumed that $\delta$ is zero-mean in (2).

We evaluate the performance of the MMSE predictor (4), which provides a fundamental limit on the predictability of $Y$ from $X$. Conditioned on $X$, the MMSE is given by the conditional variance $\text{var}(Y|X)$. Taking expectations with respect to $X$ and denoting by $M_W$ the joint moment-generating function of $W$, we have
\[
\text{MMSE} = E[\text{var}(Y|X)] = c_0^2 \sigma^2_M M_W(2\beta), \tag{5}
\]
where $\beta = (\beta_1, \ldots, \beta_p)$. For comparison, the variance of $Y$ can be determined from the law of total variance [9] $\text{var}(Y) = E[\text{var}(Y|X)] + \text{var}(E(Y|X))$.

(4) is the conditional mean $E[Y|X]$. Under the multiplicative model (1), this is given by
\[
E(Y|X) = c_0 \mu_0 \prod_{i=1}^{p} X_{i}^{\beta_i}. \tag{4}
\]

The MMSE predictor is thus scaled by the mean $\mu_0$, in addition to $c_0$, as has been noted e.g. in [1, 2, 4]. In the conventional OLS method discussed in Section 3.1, $c_0$ and $\mu_0$ are estimated separately, whereas in the NLS method in Section 3.2, a single combined scale factor is estimated. An alternative possibility is to assume that $\mu_0 = 1$, in effect absorbing $\mu_0$ into $c_0$, but in that case it cannot also be assumed that $\delta$ is zero-mean in (2).

We evaluate the performance of the MMSE predictor (4), which provides a fundamental limit on the predictability of $Y$ from $X$. Conditioned on $X$, the MMSE is given by the conditional variance $\text{var}(Y|X)$. Taking expectations with respect to $X$ and denoting by $M_W$ the joint moment-generating function of $W$, we have
\[
\text{MMSE} = E[\text{var}(Y|X)] = c_0^2 \sigma^2_M M_W(2\beta), \tag{5}
\]
where $\beta = (\beta_1, \ldots, \beta_p)$. For comparison, the variance of $Y$ can be determined from the law of total variance [9] $\text{var}(Y) = E[\text{var}(Y|X)] + \text{var}(E(Y|X))$.

(4) is the conditional mean $E[Y|X]$. Under the multiplicative model (1), this is given by
\[
E(Y|X) = c_0 \mu_0 \prod_{i=1}^{p} X_{i}^{\beta_i}. \tag{4}
\]

In the multiplicative setting (1), the prediction performance as measured by $\text{MSE}/\text{var}(Y)$ or $R^2$ can depend separately on the magnitudes of the “signal” component $X$ and the noise $\epsilon$, in contrast to the more familiar case of linear prediction where the dependence is only through the signal-to-noise ratio (SNR). Furthermore, $R^2$ can be bounded away from its maximum value of 1 depending on the noise magnitude, suggesting that multiplicative prediction is inherently difficult. To illustrate these phenomena, we assume that $\delta$ is normal, i.e. $\delta \sim \mathcal{N}(0, \sigma^2_\delta)$ and $\epsilon$ is log-normal, and $W$ is multivariate normal with covariance matrix $\Sigma_W$. Then the moments of the log-normal distribution and $M_W(\beta) = \exp((1/2)\beta^T \Sigma_W \beta)$ to evaluate (5) and (6), we obtain
\[
R^2 = \frac{\exp(\beta^T \Sigma_W \beta) - 1}{\exp(\beta^T \Sigma_W \beta + \sigma^2_\delta) - 1}.
\]
As the signal energy $\beta^T \Sigma_W \beta$ increases, $R^2$ converges quickly to an upper bound of $e^{-\sigma^2_\delta} < 1$. Thus performance is limited by the noise variance $\sigma^2_\delta$ and can be quite poor even for moderate $\sigma^2_\delta$. In contrast, for the linear problem of predicting $Z$ from $W$ in (2), the optimal coefficient of determination is
\[
R^2 = \frac{\beta^T \Sigma_W \beta}{\beta^T \Sigma_W \beta + \sigma^2_\delta},
\]
which is a function only of the SNR $\beta^T \Sigma_W \beta/\sigma^2_\delta$ and converges to 1 as the SNR increases.

3. REGRESSION METHODS

In most applications, the parameters $c_0$, $\beta$, and $\mu_0$ in the multiplicative model (1) are not known and must be estimated from sample data. We refer to this process as multiplicative regression. In Section 3.1, we discuss the most popular regression method in which OLS is applied after a logarithmic transformation. In Section 3.2, the alternative method of NLS on the original scale is discussed. In Section 3.3, a combined method is proposed that minimizes the NLS cost while constrained by the logarithmic-scale OLS solution.

3.1. Logarithmic-scale ordinary least squares

As seen in (2), the multiplicative model is transformed into a standard linear model upon taking the logarithm. Therefore the parameters $\beta_0$ and $\beta$ may be estimated using linear regression. Let $z$ represent a vector of $n$ i.i.d. samples of $Z$, and $W$ an $n \times (p + 1)$ matrix where the first column consists of ones and the remaining $p$ columns consist of $n$ i.i.d. samples each of $W_i$, $i = 1, \ldots, p$. Then the OLS estimate of $\beta = (\beta_0, \beta)$ is
\[
\hat{\beta}_\text{OLS} = (W^T W)^{-1} W^T z. \tag{7}
\]
Under the assumption in Section 2 that $E[\delta|W] = 0$, $\hat{\beta}_\text{OLS}$ is an unbiased and consistent estimator of $\beta$. If we assume further that $\delta$ is normal, then $\hat{\beta}_\text{OLS}$ corresponds to the maximum likelihood estimator and is also efficient.

Given a new sample of $W$, a “plug-in” prediction of the corresponding value of $Y$ may be obtained by substituting the OLS estimate $\hat{\beta}_\text{OLS}$ into the optimal functional form (4). However, as seen from (4) and pointed out in [4], neglecting the additional factor $\mu_0$ may lead to a large bias. To estimate $\mu_0$, the “smearing” estimator of [4] is commonly used. Defining the linear regression residual
\[
r = z - W \hat{\beta}_\text{OLS} = \left(I - W (W^T W)^{-1} W^T \right) \delta,
\]
where $\delta$ is an $n \times 1$ vector of samples of $\delta$, the smearing estimate is the empirical mean of the exponentiated residuals,
\[
\hat{\mu}_0 = \frac{1}{n} \sum_{i=1}^{n} e^{r_i}. \tag{8}
\]
This estimator is consistent for any distribution of $\epsilon$ under some regularity conditions [6]. Representing the new sample of $W$ as the row vector $w' = [1 \ w'_1 \ \ldots \ w'_p]$, the transformed OLS predictor can be expressed as
\[
\hat{Y}_\text{OLS} = \hat{\mu}_0 \exp (w' \hat{\beta}_\text{OLS}). \tag{9}
\]

3.2. Original-scale nonlinear least squares

As an alternative to logarithmic transformation and back-transformation, we may consider working on the original scale directly. Based on (4), we construct a multiplicative predictor
\[
\hat{Y}_{\text{NLS}} = c_{\text{NLS}} \prod_{i=1}^{p} X_{i}^{\hat{\beta}_{\text{NLS}}}, \tag{10}
\]
parameterized by $c_{\text{NLS}}$ and $\hat{\beta}_{\text{NLS}}$. These parameters are determined from sample data $y_1, \ldots, y_n$ and $x_{1i}, \ldots, x_{ni}$, $i = 1, \ldots, p$, by
solving the following NLS problem on the original scale:

\[
(c_{\text{NLS}}, \beta_{\text{NLS}}) = \arg \min_{c, \beta} \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\hat{c}}{\sum x_{i}^{2}} - y_{i} \right)^{2}.
\]  

The objective function in (11) is the empirical version of the MSE (3), and hence NLS may be seen as a more direct approach to MMSE prediction. A similar method has been suggested in [1]. As the number of samples \( n \) increases to infinity, the empirical MSE converges almost surely to the population MSE. It can be shown therefore that \( c_{\text{NLS}} \) converges to the product of \( c_{0} \) and \( \mu_{c} \), thus avoiding the need to estimate these factors separately as in the transformed approach in Section 3.1.

Unfortunately, for finite sample sizes encountered in practice, we have observed and report in Section 4.1 that the NLS estimates can be highly variable and can deviate greatly from the true parameters. Since these estimates are used in the exponent of the predictor (10), the deviations are exacerbated and the predictor can perform extremely poorly. In the next subsection, we present a constrained modification of (11) to address these problems.

3.3. Original-scale constrained least squares

Our proposed CLS method is based on confidence regions constructed around the OLS estimates in Section 3.1. For the OLS estimator of \( \beta \) in (7), under the assumption that \( \delta \sim N(0, \sigma_{\delta}^{2}) \) or the sample size \( n \) is large enough to invoke the central limit theorem, it can be seen that \( \hat{\beta}_{\text{OLS}} \sim N \left( \hat{\beta}, \sigma_{\delta}^{2} (W^{T} W)^{-1} \right) \). The usual elliptical confidence region can then be used, i.e., \( \hat{\beta} \) satisfies

\[
\left( \beta - \hat{\beta}_{\text{OLS}} \right)^{T} \frac{W^{T} W}{n} \left( \beta - \hat{\beta}_{\text{OLS}} \right) \leq \frac{(p + 1)\sigma_{\delta}^{2}}{n} F^{-1}(1 - \alpha)
\]

with probability \( 1 - \alpha \), where \( \sigma_{\delta}^{2} = \| \delta \|^{2} / (n - p - 1) \) is the standard estimate of \( \sigma_{\delta}^{2} \) and \( F^{-1}(1 - \alpha) \) is the quantile function of the \( F \)-distribution with \( (p + 1, n - p - 1) \) degrees of freedom.

For the smearing estimator (8) of \( \mu_{c} \), when \( n \) is large we may assume that \( \hat{\mu}_{c}^{\delta} \) is approximately normally distributed and show that its mean is \( \mu_{c} \) and its variance is bounded from above by \( \sigma_{\delta}^{2} / n \). The variance \( \sigma_{\delta}^{2} \) can be estimated in turn using the sample variance of the exponentiated residuals,

\[
\hat{\sigma}_{\delta}^{2} = \frac{1}{n - 1} \sum_{i=1}^{n} \left( e^{x_{i}} - \hat{\mu}_{c}^{\delta} \right)^{2}.
\]

Given \( \hat{\mu}_{c}^{\delta} \) and \( \hat{\sigma}_{\delta}^{2} \), we construct a level-(1 - \( \alpha \)) confidence interval for \( \mu_{c} \), bounded by

\[
\hat{\mu}_{c}^{\delta} \pm \frac{\hat{\sigma}_{\delta}}{\sqrt{n}} T^{-1} \left( 1 - \frac{\alpha}{2} \right),
\]

where \( T^{-1} \) is the quantile function for the \( t \)-distribution with \( n - 1 \) degrees of freedom.

Based on the confidence regions defined by (12) and (13), we formulate the following constrained version of problem (11):

\[
\begin{align*}
\left( \hat{\mu}_{c}, \beta_{\text{CLS}} \right) &= \arg \min \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\hat{c}}{\sum x_{i}^{2}} - y_{i} \right)^{2} \quad \text{s.t.} \\
\left( \hat{\beta} - \beta_{\text{OLS}} \right)^{T} \frac{W^{T} W}{n} \left( \hat{\beta} - \beta_{\text{OLS}} \right) &\leq \frac{(p + 1)\sigma_{\delta}^{2}}{n} \left( \mu_{c}^{\delta} - \hat{\mu}_{c}^{\delta} \right) \leq a_{2} \frac{\sigma_{\delta}}{\sqrt{n}}.
\end{align*}
\]  

The variables \( \hat{\mu} \) and \( \hat{\beta} \) are initialized to the smearing estimate \( \hat{\mu}_{c}^{\delta} \) and the OLS estimate \( \beta_{\text{OLS}} \) respectively. The constraints, parameterized by \( a_{1}, a_{2} \geq 0 \), control the subsequent deviations, with the first constraint in particular helping to stabilize the solution of (14). The parameters \( a_{1} \) and \( a_{2} \) take the place of the quantile functions in (12), (13) and can be chosen to correspond to a confidence level \( \alpha \) or to optimize prediction performance (e.g., through cross-validation).

4. NUMERICAL RESULTS

4.1. Synthetic data

The three regression methods described in Section 3 are evaluated on synthetic datasets to compare their prediction accuracies under different conditions. Each experiment below consists of 10⁴ trials. In each trial, the regression methods are applied to \( n = 1000 \) training samples \( (y_{i}, x_{i})_{i=1}^{n} \) to estimate the model parameters. We use MATLAB’s NLS solver lsgnonlin and constrained optimization solver fmincon for (11) and (14) respectively; the solutions appear to be invariant to the choice of initial values. The trained models are then used to predict \( Y \) for 10⁴ test samples of \( X \). The MSE (3) is approximated by the average over all test samples and trials.

In the first experiment, data is generated according to the multiplicative model (1) with \( \delta \sim N(0, \sigma_{\delta}^{2}) \), \( W \sim N(0, 1) \) with \( p = 3 \), \( \beta_{0} = 1 \), and \( \beta = (1, 1, 1) \). This yields log-normal distributions for \( c, X, \) and \( Y \). As noted in Section 3.1, when \( \delta \) is normal, log-scale OLS is an efficient estimator of \( \hat{\beta} \). Our results confirm that this relative efficiency carries over into prediction MSE: As the constraint tolerance \( a_{2} \) in (14) increases from zero, the MSE of CLS increases gradually (the plot is omitted due to space constraints). At the other extreme, NLS results in highly variable parameter estimates, leading to prediction MSE values that are many orders of magnitude higher in both a fraction of trials as well as the overall average.

The comparison between OLS and CLS changes when additional terms depending on \( X \) are present but not modeled, a common scenario with real-world data. We generate data with a quadratic term \( 0.1W_{2}^{2} \) added to the linear representation (2) but continue to use the nominal models (1), (2) to estimate parameters and make predictions. Fig. 1 compares the MSE of OLS and CLS as a function of \( a_{1} \) and the noise variance \( \sigma_{\delta}^{2} \). The tolerance \( a_{2} \) is chosen so that \( a_{1} \) and \( a_{2} \) correspond to the same confidence level in (12) and (13). As Fig. 1 illustrates, CLS is better able to handle model mismatch when \( a_{1} \) is chosen appropriately, specifically by allowing \( \beta_{0} \) to increase to partially capture the effect of the quadratic term. The improvement is largest at small noise variances \( \sigma_{\delta}^{2} \) that are
more comparable to the magnitude of the quadratic term, diminishing as the noise becomes dominant. NLS remains plagued by poor parameter estimates and extremely high prediction MSE.

We return to the standard model (1) and consider independent gamma distributions with shape parameters \( k_x = 1 \) for \( X_i, i = 1, \ldots, p \) (i.e., exponential distributions) and \( k_\epsilon \) for \( \epsilon \). All scale parameters are normalized to 1 as they simply scale the results. We set \( p = 5, \beta = 1 \) and \( \beta = (1, 1, 1, 2, 3) \). In this case, both OLS and NLS produce grossly inaccurate parameter estimates in a fraction of trials (around 5% for OLS) and correspondingly extreme prediction MSE. For OLS, the problem lies in the smearing estimate \( \hat{\mu}_x \) of the transformation factor. In contrast for CLS, when \( a_1 \geq 2 \) the constraints in (14) become loose enough to fully correct the errors from OLS. Based on the coarse logarithmic search over \( a_1 \) that was performed, the lowest MSE is attained at \( a_1 = 2 \) and is comparable to the MMSE (5) under exact knowledge of the model parameters: for \( k_x = 1/2, 1, 2, 4, 8 \) the MSE ratios are respectively 1.12, 1.25, 1.07, 1.01, 1.24. CLS also outperforms OLS in terms of median MSE over trials with differences on the order of 1%–10%.

4.2. Health care usage data

We analyze the performance of original-scale OLS (Orig-OLS, i.e., linear regression), log-transformed OLS (OLS), NLS, and CLS on real-world doctor visit data obtained from the 1992 National Health Institute Survey [2]. The records of adults between ages 25 and 64 who had at least one doctor visit in the past 12 months were chosen as observations. The predictors considered were: gender, age in years, race (white/other), marital status (married/not), years of completed schooling, and health status (excellent, very good, good). Variables other than age and education were coded as binary. The 44,085 observations in the data were randomly partitioned into training and test sets, and the mean numbers of doctor visits were 5.95 and 5.90 respectively for these partitions. The corresponding standard deviations were 13.94 and 13.25. The constraint tolerance \( a_1 \) for CLS was determined using 5-fold cross-validation.

From the performances reported in Table 1, it can be seen that NLS and CLS are better in terms of \( R^2 \) compared to the Orig-OLS and OLS approaches. CLS has slightly higher bias but lower variance compared to NLS in the training as well as the test sets. It should be noted that with health care data poor fits and low prediction accuracies are a rule rather than an exception because of the unavailability of many explanatory variables and model mismatch even for the variables that are available [10]. Thus the improvement of NLS and CLS may be quite significant relative to the (unknown) best performance that could be achieved with this data. It is also surprising in light of the synthetic experiments in Section 4.1 that NLS performs on par with CLS in these more difficult conditions. We have found a similar pattern with other health care and capital market research data but further investigation is needed.

### Table 1. Performance of regression methods with NHIS doctor visit data. The standard deviation and relative bias (%) of the predictions, and the coefficient of determination (%) are provided.

<table>
<thead>
<tr>
<th>Method</th>
<th>Std. dev.</th>
<th>Rel. Bias (%)</th>
<th>( R^2 ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orig-OLS</td>
<td>3.50</td>
<td>0</td>
<td>6.29</td>
</tr>
<tr>
<td>OLS</td>
<td>2.75</td>
<td>-3.55</td>
<td>5.88</td>
</tr>
<tr>
<td>NLS</td>
<td>3.55</td>
<td>-0.15</td>
<td>6.45</td>
</tr>
<tr>
<td>CLS</td>
<td>3.42</td>
<td>1.10</td>
<td>6.42</td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orig-OLS</td>
<td>3.50</td>
<td>0.59</td>
<td>7.01</td>
</tr>
<tr>
<td>OLS</td>
<td>2.76</td>
<td>-2.82</td>
<td>6.62</td>
</tr>
<tr>
<td>NLS</td>
<td>3.54</td>
<td>0.37</td>
<td>7.10</td>
</tr>
<tr>
<td>CLS</td>
<td>3.41</td>
<td>1.70</td>
<td>7.10</td>
</tr>
</tbody>
</table>

5. REFERENCES


